W5. Logistic Regression

Phuong Anh Trinh - 11200417

October 2022

1 Exercise 1

$L = -y.log(\hat{y} - (1-y).log(1-\hat{y})$

$\Rightarrow \dfrac{\partial L}{\partial w} = \dfrac{\partial L}{\partial \hat{y}}.\dfrac{\partial \hat{y}}{\partial z}.\dfrac{\partial z}{\partial w}$ (chain rule) (1)

+) $\dfrac{\partial L}{\partial \hat{y}} = -\dfrac{y}{y} + \dfrac{1-y}{1-\hat{y}} = \dfrac{-y + y.\hat{y} + \hat{y} - y.\hat{y}}{\hat{y}(1-\hat{y})} = \dfrac{\hat{y}-y}{\hat{y}(1-\hat{y})}$

+) $\hat{y} = sigmoid(z) = \dfrac{1}{1+e^{-z}}$ with $z = W^T X + b$

$\Rightarrow \dfrac{\partial \hat{y}}{\partial z} = \hat{y}(1-\hat{y})$

+) $z = W^T X + b = w_1x_1 + w_2x_2 + ... + w_nx_n$

$\Rightarrow \dfrac{\partial z}{\partial w} = x_i$

Replace into (1):

$\dfrac{\partial L}{\partial w} = \dfrac{\hat{y}-y}{\hat{y}(1-\hat{y})}.y(1\hat{-}\hat{y}).x_i$

$= (\hat{y}-y).x_i$

2 Exercise 5

a) Loss binary cross entropy is convex function with W

$L = -y.log(\hat{y} - (1-y).log(1-\hat{y})$

$\Rightarrow \dfrac{\partial L}{\partial w} = (\hat{y}-y).x$

$\Rightarrow \dfrac{\partial^2 L}{\partial w^2} = x^2\hat{y}(1-\hat{y})$

Because $x^2\hat{y}(1-\hat{y}) \geq 0, \hat{y} \in [0,1]$

$\Rightarrow$ L is convex

b) Loss mean square error is non-convex function with W

$J = (y-\hat{y})^2$

We have $\hat{y} = \dfrac{1}{1+e^{W^T X + b}}$

$\Rightarrow \dfrac{\partial \hat{y}}{\partial w} = x.\hat{y}(1-\hat{y})$

$\dfrac{\partial J}{\partial w} = \dfrac{\partial J}{\partial \hat{y}}.\dfrac{\partial \hat{y}}{\partial w}$
$= -2(y - \hat{y}).x(1 - y)\hat{y}$
$= -2x(y - \hat{y})(\hat{y} - \hat{y}^2)$
$= -2x(y\hat{y} - y\hat{y}^2 - \hat{y}^2 + \hat{y}^3)$
$\Rightarrow \dfrac{\partial^2 J}{\partial w^2} = -2x[xy\hat{y}(1 - \hat{y}) - 2xy\hat{y}\hat{y}(1 - \hat{y}) - 2x\hat{y}\hat{y}(1 - \hat{y}) + 3x\hat{y}^2\hat{y}(1 - \hat{y})]$
$= -2x^2\hat{y}(1 - \hat{y})(y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2)$
Because $x^2\hat{y}(1 - \hat{y}) \geq 0, \hat{y} \in [0, 1]$
$\Rightarrow$ Consider $f(\hat{y}) = -2x^2\hat{y}(1 - \hat{y})(y - 2y\hat{y} - 2\hat{y} + 3\hat{y}^2)$
We know that y takes only 2 values 0 and 1
+) When $y = 0, f(\hat{y}) = -6\hat{y}^2 + 4\hat{y}$
$\Rightarrow f(\hat{y}) \leq 0$ if $\hat{y} \in [0; \dfrac{1}{3}]$
+) When $y = 1, f(\hat{y}) = -6\hat{y}^2 + 8\hat{y} - 2$
$\Rightarrow f(\hat{y}) \leq 0$ if $\hat{y} \in [\dfrac{2}{3}; 1]$
$\Rightarrow \exists \hat{y} : f(\hat{y}) < 0$
$\Rightarrow$ J is non-convex